

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

## On the chemical diversity of the prebiotic ocean of early Earth

### This is the author's manuscript

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/1508625> since 2016-06-23T10:07:53Z

*Published version:*

DOI:<http://dx.doi.org/10.1017/S1473550414000688>

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

# On the chemical diversity of the prebiotic ocean of early Earth

Carlo Canepa<sup>1\*</sup>

## Abstract

This work investigates the consequences on the diverse number of chemical species in a pre-biotic terrestrial aqueous environment endowed with an amino acid source induced by the spontaneous build-up of catalytically active polypeptides from amino acid monomers. The assumed probability that a randomly-formed polypeptide exhibits catalytic properties is dependent on constraining both the chemical identity and the position of a fraction of the amino acid residues. Within this hypothesis, and using values of the average length  $n$  of the catalytic polypeptides about one half of the present-day enzymes, the stationary-state concentration of the catalytically active polypeptides is  $\approx 10^{-30} - 10^{-19}$  M, and the ratio of the concentration of a product of a catalytic process to the initial concentration of the corresponding substrate is predicted to be  $\approx 10^{-6} - 10^5$ . Matching the mean life of each catalytic polypeptide to the mean life of its substrate ( $\lambda \approx \omega$ ) is only possible by significantly rising the intensity of the source of the amino acid monomers. Under these hypothetical optimal conditions, the mean lives of the catalytic polypeptides and their substrates have values  $\omega^{-1} \approx \lambda^{-1} \approx 10$  yr and the asymptotic concentration of each product is of the same order of magnitude as the concentration of the substrate. In all cases the catalytic efficiency necessary to form the active peptides takes the typical values of present-day enzymes.

Keywords: prebiotic chemistry; enzymes

<sup>1</sup> Dipartimento di Chimica, Università di Torino, Via Pietro Giuria 7, 10125 Torino, Italy

\* Corresponding author: carlo.canepa@unito.it

## 1. Introduction

Nucleic acids and proteins currently form the basis for the informational content in living organisms. The formation of these macromolecules is highly regulated today, and the question of how the molecules themselves and the complex mechanism of regulation could have originated from the much simpler inorganic and organic compounds present in the early solar system is still open (Bada 2004). A lot of effort has been expended on the spontaneous formation of the precursors of informational molecules (nucleotides and amino acids) from the compounds present in the early atmosphere (Rode 1999), meteorites and comets (Muñoz Caro *et al.* 2002), but very little is known about the processes at the base of the formation of functional macromolecules like proteins from their monomeric constituents. The primitive conditions on Earth and the various routes for the formation of amino acids and peptides were reviewed by Rode (1999), who also provided a rationale for favoring polypeptides over RNA as the first replicating species, based on their higher chemical stability. Another open question concerning RNA and proteins is the origin of their informational content. Nucleic acids contain the information necessary to build proteins, and in turn proteins derive their structural or catalytic properties from their amino acid sequence and structure, so it has to be determined how the correct sequences and structures developed in primitive polypeptides, which could not have formed by decoding any information in pre-existing nucleic acids. The accepted paradigm is by way of selection from a large number of random combinations of primary structures governed by chance. However, there are unresolved questions regarding the practical realization of this process. First, it has to be determined how the overall number of possible polypeptides and the number of actually catalytically active polypeptides compare and how both of these numbers relate to the mean time of the reactivity for peptide bond formation and the available time ( $\sim 0.5$  Gyr) between the cooling of the Earth's crust and the first fossilized unicellular organisms (Rode 1999). Second, given the spontaneous formation of catalytic polypeptides, we need to know whether their concentrations were sufficient to actually act on various substrates and form species which would not otherwise form spontaneously in the prebiotic environment. It is logically necessary that the species generated by the catalytic process would themselves act as new substrates, thus enhancing the chemical diversity of the early Earth with the other constituents of living organisms, like lipids and sugars. In order to be able to answer some of these questions quantitatively, it is necessary to build a model of the process of polymerization under reasonable assumptions regarding the physicochemical conditions of the prebiotic environment, being aware that too many assumptions would oversimplify the problem making the ensuing answers unrealistic, while, on the other hand, taking into consideration too many factors would quickly make the model mathematically and computationally intractable.

After the previous quantitative analysis (Canepa 2013) of the catalytic requirement of the (isothermal) prebiotic aqueous environment that would allow the formation of polypeptides with the minimum length required to exhibit catalytic properties, this work examines the consequences of the spontaneous formation of a mixture of polypeptides of various lengths on the diversity of the prebiotic oceanic chemistry on the early Earth. Following the formation of proto-enzymes in the presence of chemical species acting as substrates, we may envision the development of a more complex chemistry until a kinetic cycle of reactions involving self-replicating species was reached. We assume the presence of a continuous source of amino acids that would keep the monomer concentration constant. It is an established fact that micrometeorites, interplanetary dust particles (IDP), and carbonaceous chondrites presently deliver amino acids to the Earth (Anders 1989), and many experiments have shown that amino acids could have formed in the early atmosphere subjected to various energy sources. The model developed in this work assumes the presence of an ideal amino acid source, it is immaterial if the source is endogenous or external to the planet Earth.

## 2. Model and simulation

We briefly summarize in this section the key results of a previous investigation (Canepa 2013) that are needed before we may proceed to analyze the concentration of the chemical species formed by the catalytic activity of the polypeptides. These results will be used in the next section, and we refer the reader to Canepa (2013) for further details. An active enzyme with  $n$  amino acid residues could have formed spontaneously from its monomeric components in a catalytic environment given an appropriate source of amino acids. The fraction  $\alpha$  of constrained amino acid (De Ley 1968), the set  $p$  of available amino acids, the oceanic volume  $V$ , and the probability  $p^{-\alpha n}$  that a particular polypeptide of length  $n$  and concentration  $x_n$  would exhibit catalytic properties are related by

$$f = x_n V N_A p^{-\alpha n}, \quad (2.1)$$

where  $N_A$  is the Avogadro number, and  $f$  can be thought of as the number of molecules of catalytic polypeptides in the whole volume or the ratio between the total number of molecules of length  $n$  and the number of possible primary structures. Setting  $f \cong 1$ , we determined the minimal requirements of the catalytic environment and the amino acid source that allowed the formation of a catalytic polypeptide, regardless of its concentration.

In the presence of an amino acid source with the ability to keep the concentration of monomers  $x_1$  constant, the formation of the peptide with  $n$  residues is governed by the equation

$$\dot{x}_n = \eta x_1 x_{n-1} - \omega x_n, \quad (2.2)$$

with solution

$$x_n = \zeta^{n-1} x_1 [1 - Q(n-1, \omega t)], \quad (2.3)$$

where  $Q(n-1, \omega t)$  is the ratio between the incomplete and complete gamma function of Euler

$$Q(n-1, \omega t) = \frac{\Gamma(n-1, \omega t)}{\Gamma(n-1)}. \quad (2.4)$$

We define  $\zeta = (\omega - \omega_h)/\omega$  as the ratio between the rate of peptide formation over the overall reaction rate (peptide formation plus hydrolysis). The interval of variation of the parameter  $\zeta$  describes the various effects that the environment might have had on the rate of formation of the peptide bond in polypeptides. The value  $\zeta = 1/2$  implies equal rates for the formation of the peptide bond and its hydrolysis. On the other hand, values of  $\zeta$  near unity describe very effective catalysis by either chemical species in solution or mineral surfaces. The proper frequency of the reactive system  $\omega = \eta x_1 + \eta_h x_w$  is given in terms of the bimolecular rate coefficient for peptide bond formation ( $\eta$ ) and the rate coefficient for hydrolysis ( $\eta_h$ ). The key parameter  $\zeta$  expresses the catalytic efficiency of the polymerization and it will represent the independent variable in the following discussion. In this context, both the frequency for hydrolysis  $\omega_h = \eta_h x_w$  and the frequency for spontaneous amino acid polymerization  $\omega$  determine the time scale of the problem. Both depend on the reaction rate coefficients, which in turn are functions of temperature. The experimental value for the hydrolysis of glycyl-glycine ( $6.3 \times 10^{-11} \text{ s}^{-1}$ ) (Radzicka & Wolfenden 1996) will be used in all the calculations, thus neglecting the contributions of the temperature gradient of the ocean, the lowered pH of the surface waters, and the degradative effect of radiolysis and photolysis on the polypeptides. Since we consider the rate coefficients for amino acid

hydrolysis to be constant, we will obtain upper bounds to the final concentrations of polypeptides and products of catalysis.

It was shown that the inflection point of  $Q$  at  $\omega t = n - 2$  is reached in a time span lower than  $\approx 10^5$  yr, so that it may be safely assumed that, for periods of time of the order of 1 Gyr, the dynamics of the possible formation of polypeptides of length  $n$  in the presence of catalysis affords a geometric progression with the terms

$$x_n = \zeta^{n-1} x_1 \quad (2.5)$$

for the stationary concentration  $x_n$  of polypeptides.

This work concerns the necessary conditions allowing the formation of significant amounts of products of the chemical transformations of pre-existing substrates promoted by the polypeptides through their own catalytic properties. Consequently, we must determine the required concentration of polypeptides affording  $f \gg 1$ .

**Eq.2.5** was linked to the relation between the source  $s$  of monomeric amino acids and their concentration at steady state by the relation

$$s = \omega_h x_1 \left( \frac{\zeta}{1 - \zeta} \right)^2, \quad (2.6)$$

having defined the source  $s = 4\pi R^2 \varphi / (VN_A)$  in terms of the incoming flux  $\varphi$  of amino acids. We note that a smaller volume of the sea with respect to the present-day value would have increased the value of the source. For the purpose of calculation (All calculations were performed with the program Scilab: Scilab Enterprises (2012). Scilab: Free and Open Source software for numerical computation (OS, Version 5.XX) [Software]. Available from: <http://www.scilab.org>), in this work we use  $s = 2.83 \times 10^{-20} \text{ M s}^{-1}$ , the mean value of Chyba *et al.* (1990) estimate, and  $s/\omega_h = 4.49 \times 10^{-10} \text{ M}$  for glycyl-glycine. At steady state we obtained, for the concentration of polypeptides of length  $n$ ,

$$x_n = \frac{s}{\omega_h} (1 - \zeta)^2 \zeta^{n-3}. \quad (2.7)$$

The efficiency of a catalyst promoting the condensation of the amino acids is also given by the ratio of the rate coefficient for catalyzed formation of the polypeptide to the corresponding rate coefficient for hydrolysis

$$\frac{\eta}{\eta_h} = \frac{x_w \omega_h}{s} \left( \frac{\zeta}{1 - \zeta} \right)^3. \quad (2.8)$$

In particular, we note that the concentration of polypeptides has a maximum for the value of the catalytic efficiency  $\zeta_m = (n - 3)/(n - 1)$ , corresponding to  $\eta/\eta_h = (x_w \omega_h / 8s)(n - 3)^3 = 1.41 \times 10^{16}$  for  $n = 100$ , a value attained by the most efficient of the present-day enzymes (**Figure 1**). The catalytic efficiency  $\zeta_m$  maximizing  $x_n$ , the frequency  $\omega$  expressing the reactivity of the polypeptides towards their formation or hydrolysis, and the ratio  $\eta/\eta_h$  are given in **Table 1** for different values of  $n$ . The effect of catalysis on the dynamics of the concentrations of all the polymers of different lengths was found to be essential. This investigation begins by determining the concentration of catalytic polypeptides of given lengths in steady-state conditions. The choice of the minimum value of  $n$  that allows for catalytic features is somehow speculative. Short

polypeptides, even dipeptides, have been shown to possess the ability to promote at pH = 10 the formation of the peptide bond between model systems such as ethyl- or thioesters of amino acids and amino acid derivatives such as the amide of leucine (Gorlero *et al.* 2009, Lee *et al.* 1996). However, free amino acids do not react in the same conditions, probably because the high values of pH generating the nucleophile  $\text{NH}_2$  from  $\text{NH}_3^+$  at the same time deprotonates the carboxylic group to  $\text{COO}^-$ , thus impairing the reactivity of the electrophile.

### 3. Results

We first observe that not every combination of amino acids in a given sequence corresponds to a catalytic polypeptide. In order to be an effective catalyst, a polypeptide must have specific structural properties resulting in its ability to bind a substrate and to promote its chemical transformation. This ability is summarized by the value of the ratio  $k_2/k_M$  in the Michaelis-Menten mechanism of enzyme catalysis. The concentration of catalytically active polypeptides is  $(x_n)_{cat} = x_n p^{-\alpha n}$ , where the fraction  $\alpha$  of constrained amino acid residues in specific positions in the chain implies catalytic activity. Plots of  $(x_n)_{cat}$  are shown in **Figure 2** for  $p = 10$ ,  $\alpha = 0.10$  and  $n = 60, 80, 100$ , ostensibly showing the dramatic effect of the values of these parameters on the concentration of a polypeptide of given length.

We now turn our attention to the formation of possible reaction products catalyzed by the polypeptides by considering as active catalysts all polypeptides with  $\alpha n$  amino acid residues constrained. The Michaelis-Menten reaction rate for the transformation of a generic substrate  $z$  is proportional to the product of the concentration of the catalyst  $(x_n)_{cat}$ , and  $z$ . The rate coefficient  $\lambda$  for the process can be deduced considering that, in the primordial oceanic conditions, the concentration of any substrate was small compared to typical values of the Michaelis constant  $k_M$ . The ratio  $k_2/k_M$  is an indication of the catalytic ability of a polypeptide in promoting a chemical reaction on a substrate ( $z \rightarrow u$ ), and it must not be confused with the parameter  $\zeta$ , relating to the catalytic process forming the polypeptides. We also safely assume that the concentration  $z$  of a substrate was larger than the typical steady-state concentration of the catalyst  $(x_n)_{cat}$ , and, for the sake of simplicity, assume the source of  $z$  to be of the same intensity as the source of amino acids. If the condition  $(x_n)_{cat} \ll z \ll k_M$  is satisfied, the dynamics of the substrate follows the equation

$$\dot{z} = s - \lambda z, \quad \lambda = \frac{k_2}{k_M} x_n^0 p^{-\alpha n}, \quad (3.1)$$

with solution

$$\frac{z}{z_0} = e^{-\lambda t} + \frac{s}{z_0 \lambda} (1 - e^{-\lambda t}). \quad (3.2)$$

In this context  $x_n^0$  in the expression for  $\lambda$  is the stationary concentration (**Eq.2.7**) of polypeptide of length  $n$  when it reaches catalytic capability. For the purpose of calculation, we attribute to  $z_0$  (the concentration of the substrate before the onset of catalysis) the value estimated by Pierazzo and Chyba (Thomas *et al.* 2006) for the oceanic glycine concentration from electrical discharge and cometary input 4.0 Gyr ago ( $2.4 \times 10^{-9}$  M). Using this estimate, we obtain  $s/z_0 = 1.18 \times 10^{-11} \text{ s}^{-1}$ , a value close to  $\omega_h$ . The frequencies  $s/z_0$ ,  $\lambda$ , and  $\omega$  are plotted in **Figure 3** for  $k_2/k_M = 1.0 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ ,  $p = 10$ ,  $\alpha = 0.10$ , and three different values of  $n$ . The frequency  $\omega$  of polymerization is of the order  $10^{-2} \text{ yr}^{-1}$  for most values of  $\zeta$ , except around  $\zeta \approx 1$ . The

corresponding frequency  $\lambda$  for the reactivity of the substrates is many orders of magnitude lower. Without catalysis ( $\lambda = 0$ ),  $z$  would follow the law

$$\frac{z}{z_0} = 1 + \frac{st}{z_0}, \quad (3.3)$$

and the concentration of the reaction product  $u$  is thus given by the difference between **Eq.3.3** and **Eq.3.2**

$$\frac{u}{z_0} = (1 - e^{-\lambda t}) \left(1 - \frac{s}{z_0 \lambda}\right) + \frac{st}{z_0}. \quad (3.4)$$

In all the cases where  $\lambda t$  is small ( $n > 60$ ), the concentration of the product may be approximated as  $u/z_0 \cong \lambda t(1 - \lambda t/2 + st/2z_0)$ . With the selected source and  $t = 0.5$  Gyr we have  $u/z_0 \approx 10^{-6} - 10^5$  for the above-mentioned value of  $z_0$ , even with the low values of the product-forming frequency ( $\lambda \approx 10^{-20} - 10^{-9} \text{ yr}^{-1}$ ). The effect of  $n$  on these values is dramatic. Only polypeptides with a relatively low value of  $n$  may form a non-negligible concentration of product, and it should be determined if such a low value of  $n$  is compatible with the assumed catalytic efficiency  $k_2/k_M \approx 10^3 \text{ M}^{-1}\text{s}^{-1}$ . A surface plot of **Eq.3.4** versus  $\zeta$  and the wide range of  $s$  estimated by taking into account the lunar cratering record (Hartmann *et al.* 2000) is shown in **Figure 4**. The surface plot shows the effect of values of the amino acid source significantly in excess of the estimated contribution from micrometeorites. In fact, the source is related to the flux  $\varphi$  of amino acids and the depth of the water  $h$  through the equation  $s = \varphi/h$ . It can thus be conceived that regions of shallow waters may have actually caused a locally increased value of  $s$  with a corresponding linear increase in  $\lambda$ . It is also possible that the production of more complex amino acid species from amino acids of simple structure would have increased over time the value of  $p$ . In this case, **Eq.3.1** would involve a nonlinear function  $p(u)$ . However, we know that  $p$  would not increase more than a factor of two, decreasing the value of  $\lambda$  by a factor of  $2^{-\alpha n}$ . To clarify the effect of an increasing  $p$ , we compute the state variables with the values  $p = 10$  (**Table 2**), and  $p = 20$  (**Table 3**). The catalytic efficiency is set to the value maximizing **Eq.2.7**, i.e.  $\zeta_m = (n - 3)/(n - 1)$ . Comparing the data in **Tables 2** and **3**, we see that the increase of  $p$  does not change the concentration of products obtained through catalytic polypeptides of low length. At higher values of  $n$ , however, the increase of  $p$  makes significant concentrations of the products  $u$  even more unattainable in the time frame  $t = 0.5$  Gyr. For most values of  $n$ , the proper time for reaction ( $\lambda^{-1}$ ) greatly exceeds the age of the universe ( $\tau_u$ ), indicating that the catalyzed process would be much slower than the available time for chemical evolution. Imposing the constraint  $\lambda^{-1} = \tau_u$  would imply in all cases values of  $k_2/k_M$  in excess of the limit  $10^{10} \text{ M}^{-1}\text{s}^{-1}$  for diffusion-controlled reactions. The data in **Table 2** also show that the value of  $f$  (the number of molecules of catalytic polypeptides of a given length) sharply decreases with growing  $n$ , but the limit value  $f = 1$  is never reached. The number of catalytic polypeptides is quite large, of the order  $p^{\alpha n}$ , making the number of chemical species likely limited by the number of substrates with initial concentration  $z_0$ .

For the sake of completeness, we also consider the case of a discontinued source of amino acids. In this case, an effective polypeptide from the point of view of its catalytic properties could not be re-formed in the next generation of polypeptides, its mean life being only  $\omega^{-1} = (1 - \zeta)/\omega_h \approx 10$  yr. This also limits the number of catalytic processes per unit time to  $f\omega$ . Consequently, the concentration of the catalytic species would follow the equation

$$(\dot{x}_n)_{cat} = -\omega(x_n)_{cat}, \quad (3.5)$$

and the dynamics of the substrate the equation

$$\dot{z} = s - \lambda e^{-\omega t} z, \quad (3.6)$$

with solution

$$\frac{z}{z_0} = e^{-\frac{\lambda}{\omega}(1-e^{-\omega t})} - \frac{s}{z_0 \omega} e^{\frac{\lambda}{\omega} e^{-\omega t}} \int_{a_1}^{a_2} \frac{d\xi}{\ln \xi}, \quad (3.7)$$

where  $a_1 = e^{-\lambda/\omega}$  and  $a_2 = e^{-\frac{\lambda}{\omega} e^{-\omega t}}$ . The difference between **Eq.3.3** and **Eq.3.7** gives the concentration of the reaction product  $u$

$$\frac{u}{z_0} = 1 - e^{-\frac{\lambda}{\omega}(1-e^{-\omega t})} + \frac{st}{z_0} \left( 1 + \frac{e^{\frac{\lambda}{\omega} e^{-\omega t}}}{\omega t} \int_{a_1}^{a_2} \frac{d\xi}{\ln \xi} \right). \quad (3.8)$$

Its asymptotic behavior for  $\omega t \gg 1$  is  $u_\infty/z_0 = 1 - e^{-\lambda/\omega}$ . An approximate expression for **Eq.3.8** in terms of elementary functions may be obtained for  $\lambda/\omega \ll 1$ . In this case  $\xi \approx 1$  and the logarithmic integral in **Eq.3.8** may be written as

$$\int_{a_1}^{a_2} \frac{d\xi}{\ln \xi} \cong \int_{a_1}^{a_2} \frac{d\xi}{\xi - 1} = \ln \frac{a_2 - 1}{a_1 - 1} \cong -\omega t, \quad (3.9)$$

obtaining

$$\frac{u}{z_0} = 1 - e^{-\frac{\lambda}{\omega}(1-e^{-\omega t})} + \frac{st}{z_0} \left( 1 - e^{\frac{\lambda}{\omega} e^{-\omega t}} \right). \quad (3.10)$$

In all cases under investigation  $\lambda/\omega \ll 1$ , and **Eq.3.10** may be approximated with the expression

$$\frac{u}{z_0} \cong \frac{\lambda}{\omega} \left[ 1 - \left( 1 + \frac{st}{z_0} \right) e^{-\omega t} \right], \quad (3.11)$$

with the asymptotic limit  $u_\infty/z_0 = \lambda/\omega$ . Since we have now excluded the survival of a catalytic species for more than one generation of polypeptides, if we want to attain catalysis and a significant build-up of  $u$ , we must equate the mean life for the transformation of a substrate ( $\lambda^{-1}$ ) to the mean life of the catalytic species ( $\omega^{-1}$ ), obtaining

$$(1 - \zeta)^3 \zeta^{n-3} = \frac{\omega_h^2 p^{an}}{s k_2/k_M}. \quad (3.12)$$

**Eq.3.12** does not have solutions for the values of the parameters used up to this point, and we can find solutions only by considerably rising the value of either the amino acid source or the rate coefficient  $k_2/k_M$  (or both) by a combined factor of  $9 \times 10^5$ . A plot of **Eq.3.12** and its solution  $\bar{\zeta}$ ,



obtained with the unrealistic value  $s = 2.55 \times 10^{-14} \text{ M s}^{-1}$ , is shown in **Figure 5**. Higher values of  $n$ ,  $p$ ,  $\alpha$ , or lower values of  $s$  do not afford solutions to **Eq.3.12**, that is, the half-life of transformation of a substrate would be longer than the mean life of the catalytic species  $(x_n)_{cat}$ , lowering the efficiency of the production of new chemical species. With  $\lambda = \omega$ ,  $a_1 = e^{-1}$ ,  $a_2 = e^{-e^{-\omega t}}$ , and the asymptotic condition  $\omega t \gg 1$ , the concentration of the product is  $u_\infty/z_0 = 1 - e^{-1} \cong 0.63$ .

## 4. Discussion

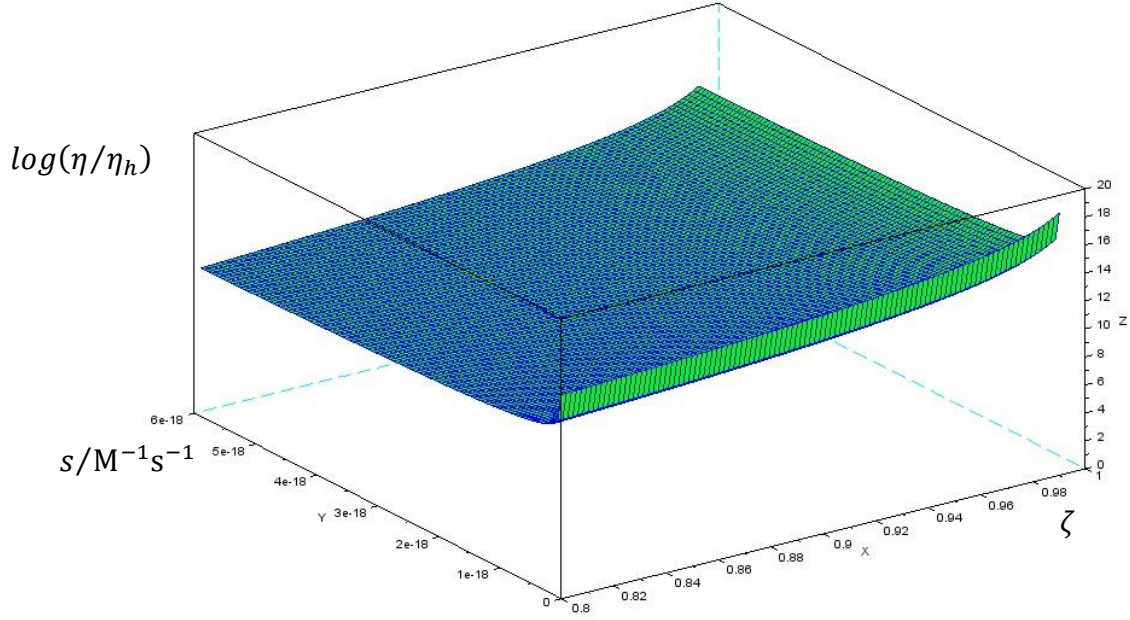
The present-day amino acid flux on Earth was estimated to be around  $3.0 \times 10^7 \text{ kg yr}^{-1}$  (Matrajt *et al.* 2004) with an amino acid content of 869 nmol/g. These figures translate into an amino acid source  $s$  of  $5.90 \times 10^{-25} \text{ M s}^{-1}$  for the present mass of the ocean ( $1.4 \times 10^{21} \text{ kg}$ ). To account for the increased bombardment rate based on the lunar cratering record of 3.8 to 4.5 Gyr ago (Hartmann *et al.* 2000), we multiply by a factor of  $10^2$  to  $10^7$  to obtain  $5.90 \times 10^{-23} < s < 5.90 \times 10^{-18} \text{ M s}^{-1}$ . A somewhat narrower interval for  $s$  is given by the estimate of Chyba (1990) of  $10^8 - 10^{10} \text{ kg yr}^{-1}$  of organic carbon with a 3% content of amino acids, that translates into a source of  $2.83 \times 10^{-21} < s < 2.83 \times 10^{-19} \text{ M s}^{-1}$  of glycine. We may also mention the flux of organic carbon in IDPs of  $3.6 \times 10^{10} \text{ kg yr}^{-1}$  that survived re-entry on the early Earth proposed by Wilson (2009). Assuming again a 3% content of amino acids, we obtain for glycine  $s = 1.02 \times 10^{-18} \text{ M s}^{-1}$ . It was shown (Canepa 2013) that, besides the intensity of the source, the length of the spontaneously-formed polypeptides depends on how many types of monomeric units are involved in the formation of the chain, the fraction  $\alpha$  of constrained amino acids, the intensity of the source  $s$ , and the catalytic efficiency for their formation  $\zeta$ . Brack (1993) reviewed the chemical species with the ability to promote the condensation of amino acid monomers (i.e. increase  $\zeta$ ) and discussed the conformational and chemical properties of the resulting polypeptides according to both the nature of the side chain in the constituent amino acids and their chirality. Among the condensing agents are carbodiimides, clays, surfactant aggregates, polyphosphates, and N-carboxyanhydrides. In conclusion, significant amounts of products might have been synthesized in a prebiotic environment containing spontaneously-formed polypeptides with catalytic properties only in the assumption that the ratio between the rate coefficients for the formation of the peptide bond and its hydrolysis exceeds  $\sim 10^{15}$ , and the fraction of conserved amino acid residues in the catalyst does not exceed 0.1. In this case, if the catalytic properties emerge for  $n \approx 60$ , the concentration of products exceeds  $z_0$  for  $k_2/k_M = 1.0 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$  and  $n \leq 100$ . A final consideration is due with regard to the parameter  $k_2/k_M$ , which in present day enzymes assumes a wide range of values. We have no means to guess how prebiotic polypeptides would position within that range, and we selected a value for the purpose of calculation that would be considered rather low for evolved enzymes. However, if the prebiotic species would exhibit a value of  $k_2/k_M$  significantly different from  $1.0 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$ ,  $\lambda$  would be correspondingly altered, with consequent dramatic changes in the concentrations  $u$  of the products of catalysis given in **Table 2**. These are the products of transformation of a generic substrate  $z$  by a catalytic polypeptide which was spontaneously formed from its constituent amino acids in an aqueous environment, itself endowed with the catalytic ability to promote the formation of the peptide bond summarized by the parameter  $\zeta$ . One should bear in mind that, while the dependence of  $\lambda$  from  $k_2/k_M$  is linear, the dependence from  $\zeta$  is not, according to **Eq.2.7**. The formation of products of catalysis with concentrations at least of the same order as the starting substrates seems thus to be unlikely under the conditions assumed in this work. A high ability of the aqueous environment to form polypeptides ( $\zeta$ ) and a substantial catalytic power of polypeptides with limited length ( $k_2/k_M$ ) are both necessary to afford significant amounts

of products. The change to this picture caused by the presence of autocatalytic species, i.e. polypeptides catalyzing the formation of the peptide bond could be the subject of forthcoming investigations.

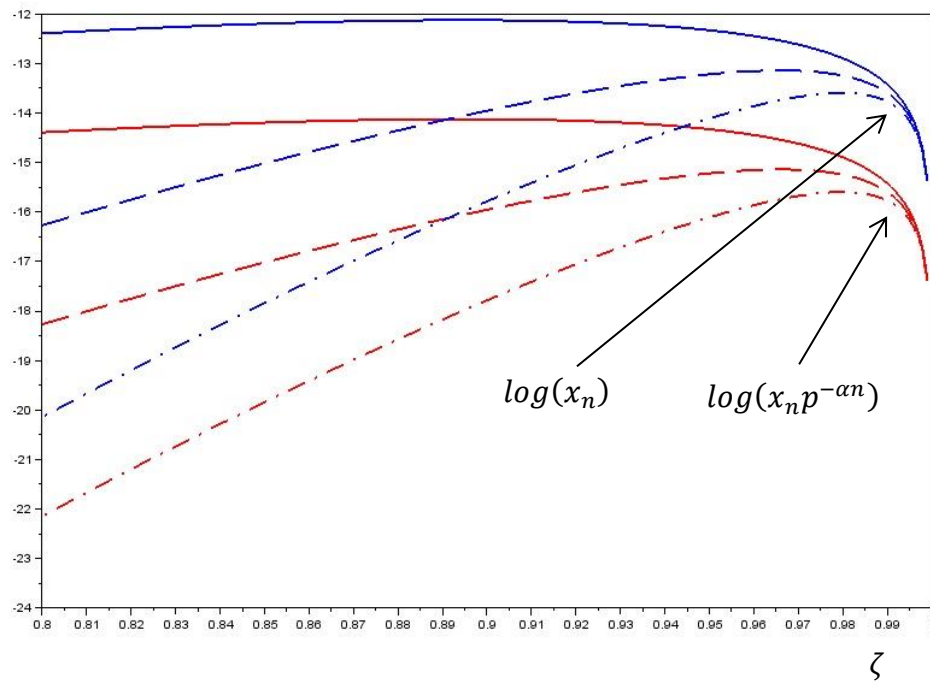
## 5. Conclusions

1. The steady-state concentration of polypeptides of length  $n$ , formed in the primordial ocean upon an incoming flux of monomeric amino acids in the presence of catalysis with efficiency  $\zeta$  is  $x_n = (s/\omega_h)(1 - \zeta)^2 \zeta^{n-3}$ .
2. Using about one half of the present-day values for the average length  $n$  of the catalytic polypeptides, the size of the amino acid set  $p$ , the fraction  $\alpha$  of constrained amino acid residues within a polypeptide, the value of the catalytic efficiency  $\zeta_m$  maximizing the spontaneous synthesis of the polypeptides, and the value  $k_2/k_M = 1.0 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$  for the catalytic ability to promote chemical transformations, we obtain  $\omega^{-1} \approx 10 \text{ yr}$  and  $\lambda^{-1} \approx 10^9 - 10^{20} \text{ yr}$  for the polypeptide mean life and the substrate mean life, respectively. The ratio of the rate coefficient for the catalyzed formation of the polypeptides to the corresponding rate coefficient for its hydrolysis exhibits the typical values of evolved enzymes.
3. For polypeptides with  $\alpha n$  constrained amino acids in specific locations along the chain, the catalyzed formation of products  $u$  enriching the chemical diversity of the prebiotic ocean gives concentrations of the order  $u(t)/z_0 \cong 10^{-6} - 10^5$  after 0.5 Gyr.
4. The value of the amino acid source  $s$  found by imposing the condition  $\lambda = \omega$  would be questionably high. Within this constraint, both the catalytic species and its substrates exhibit an average life  $\approx 10^2 \text{ yr}$ , with the asymptotic concentration of the products being  $u_\infty/z_0 = 0.63$ .

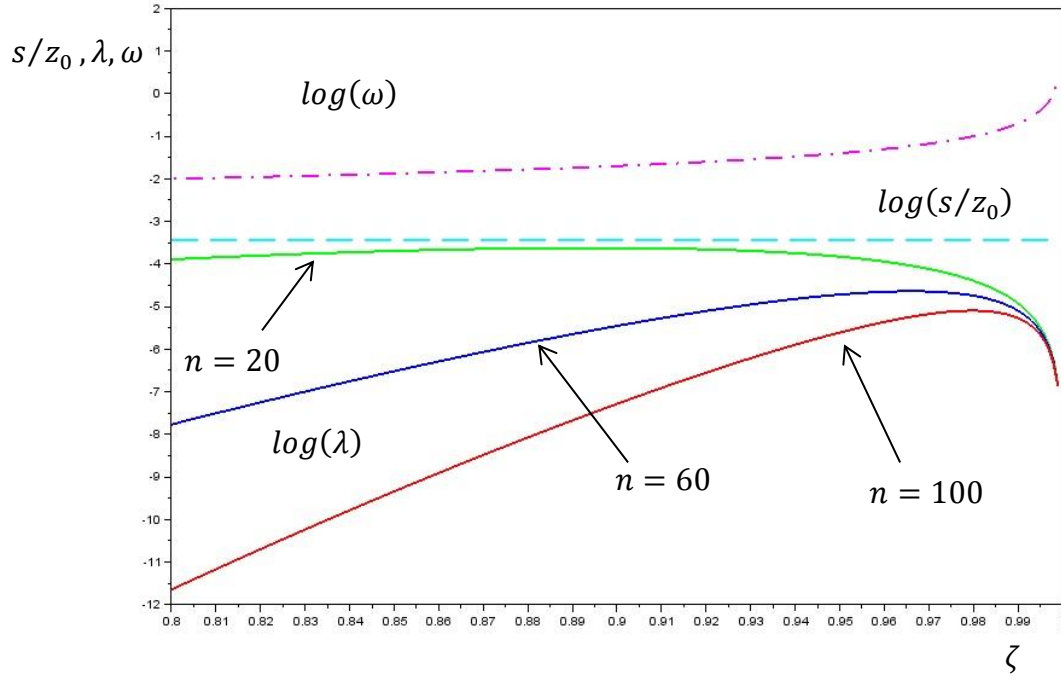
**Figure 1.** Plot of the logarithm of the ratio  $\eta/\eta_h$  (Eq.2.8) versus the catalytic efficiency  $\zeta$  and the amino acid source  $s$  for  $p = 10$  and  $\alpha = 0.10$ .



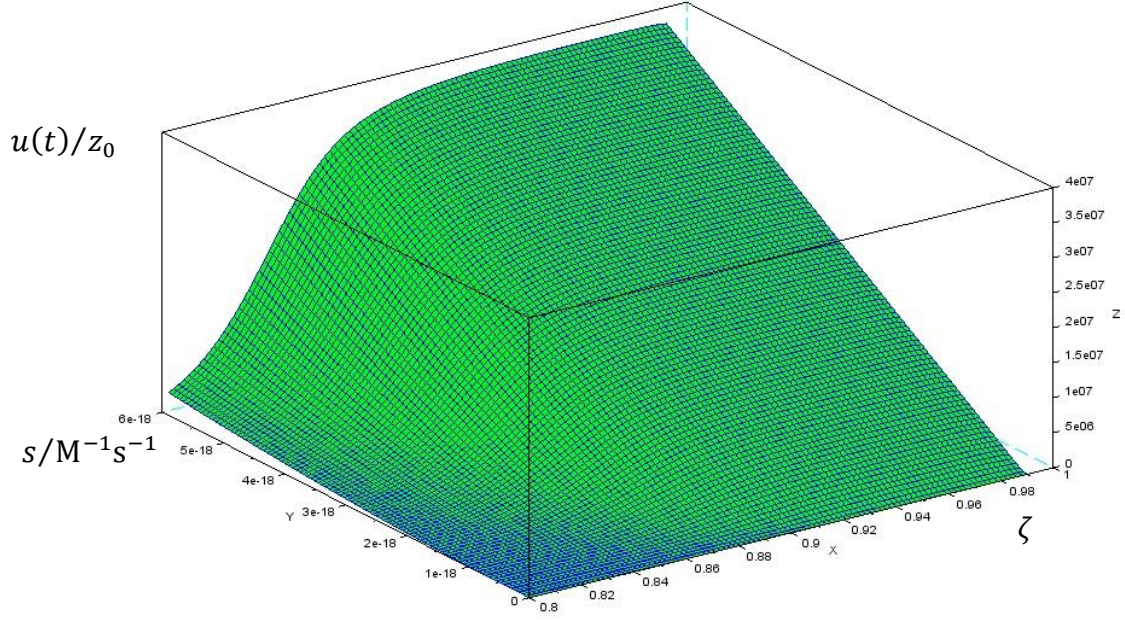
**Figure 2.** Logarithm of the molar concentration of the species  $x_n$  (blue) and of the catalytic polypeptides  $x_n p^{-\alpha n}$  (red) given by **Eq.2.7** for  $p = 10$  and  $\alpha = 0.10$ . Solid lines are for  $n = 20$ , dashed lines for  $n = 60$ , and dash-dot lines for  $n = 100$ .



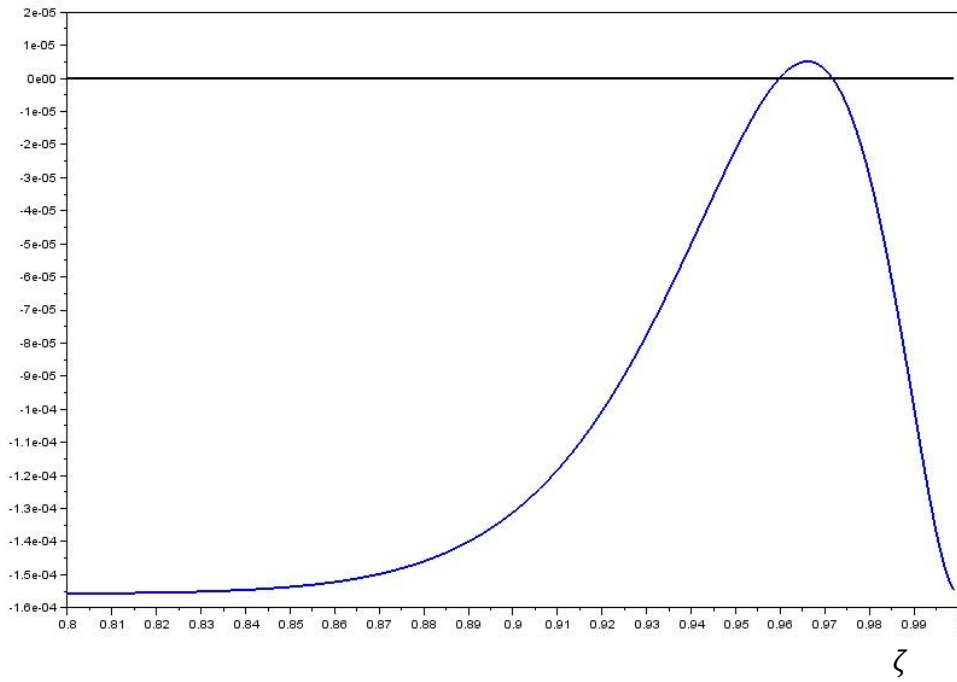
**Figure 3.** Logarithm of the frequencies (in  $\text{yr}^{-1}$ )  $s/z_0$  (dashed lines),  $\lambda$  (solid lines), and  $\omega$  (dash-dot line) versus the catalytic efficiency  $\zeta$  for  $k_2/k_M = 1.0 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$ ,  $p = 10$ , and  $\alpha = 0.10$ . The effect of different values of  $n$  is also shown:  $n = 20$  (green),  $n = 60$  (blue), and  $n = 100$  (red).



**Figure 4.** Plots of the ratio  $u(t)/z_0$  at  $t = 0.5$  Gyr versus the catalytic efficiency  $\zeta$  and the amino acid source  $s$  for  $n = 60$ ,  $k_2/k_M = 1.0 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$ ,  $p = 10$ ,  $\alpha = 0.10$ , and  $z_0 = 2.4 \times 10^{-9} \text{ M}$  given by Eq.3.4.



**Figure 5.** Plot of **Eq.3.12**, the expression for optimal condition ( $\lambda = \omega$ ) for the formation of products after turning off the amino acid source, versus the catalytic efficiency  $\zeta$  for  $n = 60$ ,  $p = 10$ ,  $\alpha = 0.1$ ,  $z_0 = 2.4 \times 10^{-9}$  M, and  $k_2/k_M = 1.0 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$ . The solution is  $\bar{\zeta} = 0.9717$ , the initial source  $s = 2.55 \times 10^{-14} \text{ M s}^{-1}$ , the molar concentration of catalytic polypeptides  $(x_n)_{cat} = 6.30 \times 10^{-14} \text{ M}$ , the frequency of formation of the polypeptides (which equals the frequency  $\lambda$  of the catalyzed reactions)  $\omega = 0.0704 \text{ yr}^{-1}$ , and the required ratio between the polymerization rate coefficient to the hydrolysis rate coefficient  $\eta/\eta_h = 5.59 \times 10^9$ .



**Table 1.** Frequency  $\omega$  of reactivity of the polypeptides and the required ratio  $\eta/\eta_h$  between the polymerization rate coefficient to the hydrolysis rate coefficient for the values of the catalytic efficiency  $\zeta_m$  maximizing the concentration of active polypeptides at various values of  $n$ ,  $p = 10$ , and  $\alpha = 0.10$ .

$n$	$\zeta_m$	$\omega/\text{yr}^{-1}$	$\eta/\eta_h$
60	0.966	$5.86 \times 10^{-2}$	$2.86 \times 10^{15}$
80	0.975	$7.86 \times 10^{-2}$	$7.07 \times 10^{15}$
100	0.980	$9.84 \times 10^{-2}$	$1.41 \times 10^{16}$
120	0.982	$1.18 \times 10^{-1}$	$2.48 \times 10^{16}$
140	0.986	$1.38 \times 10^{-1}$	$3.98 \times 10^{16}$
160	0.987	$1.58 \times 10^{-1}$	$5.98 \times 10^{16}$

**Table 2.** For various values of  $n$ ,  $p = 10$ ,  $\alpha = 0.10$ ,  $k_2/k_M = 1.0 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$ ,  $\zeta = \zeta_m$ , and  $z_0 = 2.4 \times 10^{-9} \text{ M}$  are given: the probability  $p^{-an}$ , the molar concentration of catalytic polypeptides  $(x_n)_{cat}$ , the frequency  $\lambda$  of the catalyzed reactions, the value  $\lambda t$ , and the concentration of products  $u(t)/z_0$  at  $t = 0.5 \text{ Gyr}$  given by **Eq.3.4**.

$n$	$p^{-an}$	$(x_n)_{cat}/\text{M}$	$\lambda/\text{yr}^{-1}$	$\lambda t$	$u(t)/z_0$	$\lambda/\omega$	$f$
20	$1.00 \times 10^{-2}$	$7.51 \times 10^{-15}$	$2.37 \times 10^{-4}$	$1.19 \times 10^5$	$1.86 \times 10^5$	$1.26 \times 10^{-2}$	$6.33 \times 10^{30}$
40	$1.00 \times 10^{-4}$	$1.68 \times 10^{-17}$	$5.32 \times 10^{-7}$	$2.66 \times 10^2$	$1.85 \times 10^5$	$1.37 \times 10^{-5}$	$1.42 \times 10^{28}$
60	$1.00 \times 10^{-6}$	$7.23 \times 10^{-20}$	$2.28 \times 10^{-9}$	1.14	$7.51 \times 10^4$	$3.89 \times 10^{-8}$	$6.10 \times 10^{25}$
80	$1.00 \times 10^{-8}$	$4.00 \times 10^{-22}$	$1.26 \times 10^{-11}$	$6.31 \times 10^{-3}$	$5.87 \times 10^2$	$1.61 \times 10^{-10}$	$3.37 \times 10^{23}$
100	$1.00 \times 10^{-10}$	$2.53 \times 10^{-24}$	$7.99 \times 10^{-14}$	$4.00 \times 10^{-5}$	$3.72 \times 10^0$	$8.12 \times 10^{-13}$	$2.13 \times 10^{21}$
120	$1.00 \times 10^{-12}$	$1.75 \times 10^{-26}$	$5.51 \times 10^{-16}$	$2.76 \times 10^{-7}$	$2.56 \times 10^{-2}$	$4.66 \times 10^{-15}$	$1.47 \times 10^{19}$
140	$1.00 \times 10^{-14}$	$1.28 \times 10^{-28}$	$4.03 \times 10^{-18}$	$2.02 \times 10^{-9}$	$1.87 \times 10^{-4}$	$2.92 \times 10^{-17}$	$1.08 \times 10^{17}$
160	$1.00 \times 10^{-16}$	$9.74 \times 10^{-31}$	$3.07 \times 10^{-20}$	$1.54 \times 10^{-11}$	$1.43 \times 10^{-6}$	$1.95 \times 10^{-19}$	$8.21 \times 10^{14}$

**Table 3.** For various values of  $n$ ,  $p = 20$ ,  $\alpha = 0.10$ ,  $k_2/k_M = 1.0 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$ ,  $\zeta = \zeta_m$ , and  $z_0 = 2.4 \times 10^{-9} \text{ M}$  are given: the probability  $p^{-an}$ , the molar concentration of catalytic polypeptides  $(x_n)_{cat}$ , the frequency  $\lambda$  of the catalyzed reactions, the value  $\lambda t$ , and the concentration of products  $u(t)/z_0$  at  $t = 0.5 \text{ Gyr}$  given by **Eq.3.4**.

$n$	$p^{-an}$	$(x_n)_{cat}/\text{M}$	$\lambda/\text{yr}^{-1}$	$\lambda t$	$u(t)/z_0$	$\lambda/\omega$	$f$
20	$2.50 \times 10^{-3}$	$1.88 \times 10^{-15}$	$5.93 \times 10^{-5}$	$2.96 \times 10^4$	$1.86 \times 10^5$	$3.14 \times 10^{-3}$	$1.58 \times 10^{30}$
40	$6.25 \times 10^{-6}$	$1.05 \times 10^{-18}$	$3.32 \times 10^{-8}$	$1.66 \times 10^1$	$1.75 \times 10^5$	$8.57 \times 10^{-7}$	$8.88 \times 10^{26}$
60	$1.56 \times 10^{-8}$	$1.13 \times 10^{-21}$	$3.57 \times 10^{-11}$	$1.78 \times 10^{-2}$	$1.65 \times 10^3$	$6.08 \times 10^{-10}$	$9.52 \times 10^{23}$
80	$3.91 \times 10^{-11}$	$1.56 \times 10^{-24}$	$4.93 \times 10^{-14}$	$2.46 \times 10^{-5}$	$2.29 \times 10^0$	$6.27 \times 10^{-13}$	$1.32 \times 10^{21}$
100	$9.77 \times 10^{-14}$	$2.47 \times 10^{-27}$	$7.80 \times 10^{-17}$	$3.90 \times 10^{-8}$	$3.63 \times 10^{-3}$	$7.93 \times 10^{-16}$	$2.09 \times 10^{18}$
120	$2.44 \times 10^{-16}$	$4.26 \times 10^{-30}$	$1.35 \times 10^{-19}$	$6.73 \times 10^{-11}$	$6.26 \times 10^{-6}$	$1.14 \times 10^{-18}$	$3.60 \times 10^{15}$
140	$6.10 \times 10^{-19}$	$7.79 \times 10^{-33}$	$2.46 \times 10^{-22}$	$1.23 \times 10^{-13}$	$1.14 \times 10^{-8}$	$1.78 \times 10^{-21}$	$6.57 \times 10^{12}$
160	$1.53 \times 10^{-21}$	$1.49 \times 10^{-35}$	$4.69 \times 10^{-25}$	$2.35 \times 10^{-16}$	$2.18 \times 10^{-11}$	$2.97 \times 10^{-24}$	$1.25 \times 10^{10}$



## References

- Anders, E. (1989). Pre-biotic organic matter from comets and asteroids. *Nature* **342**, 255–257.
- Bada, J. L. (2004). How life began on Earth: a status report. *Earth. Planet. Sc. Lett.* **226**, 1–15.
- Brack, A. (1993). From amino acids to prebiotic active peptides: A chemical reconstitution. *Pure Appl. Chem.* **65**, 1143–1151.
- Canepa, C. (2013). The role of catalysis on the formation of an active proto-enzyme in the prebiotic aqueous environment. *Natural Science* **5**, 549–555.
- Chyba, C. F., Thomas, P. J., Brookshaw, L., Sagan, C. (1990). Cometary delivery of organic molecules to the early earth. *Science* **249**, 366–373.
- De Ley, J. (1968). Molecular biology and bacterial phylogeny. *J. Evolution. Biol.* **2**, 103–156.
- Gorlero, M., Wieczorek, R., Adamala, K., Giorgi, A., Schininà, M. E., Stano, P., Luisi, P. L. (2009). Ser-His catalyses the formation of peptides and PNAs. *FEBS Letters* **583**, 153–156.
- Hartmann, W. K., Ryder, G., Dones, L., Grinspoon, D. (2000). The time-dependent intense bombardment of the primordial Earth/Moon system. *Origin of the Earth and Moon*. University of Arizona Press, Tucson.
- Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K., Ghadiri, M. R. (1996). A self-replicating peptide. *Nature* **382**, 525–528.
- Matrajt, G., Pizzarello, S., Taylor, S., Brownlee, D. (2004). Concentration and variability of the AIB amino acid in polar micrometeorites: Implications for the exogenous delivery of amino acids to the primitive earth. *Meteorit. Planet. Sci.* **11**, 1849–1858.
- Muñoz Caro, G. M., Meierhenrich, U. J., Schutte, W. A., Barbier, B., Arcones Segovia, A., Rosenbauer, H., Thiemann, W. H-P., Brack, A., Greenberg, J. M. (2002). Amino acids from ultraviolet irradiation of interstellar ice analogues. *Nature* **416**, 403–406.
- Radzicka, A. & Wolfenden, R. (1996). Rates of uncatalyzed peptide bond hydrolysis in neutral solution and the transition state affinities of proteases. *J. Am. Chem. Soc.* **118**, 6105–6109.
- Rode, B. M. (1999). Peptides and the origin of life. *Peptides* **20**, 773–786.
- Thomas, P. J., Hicks, R. D., Chyba, C. F., McKay, C. P. (2006). *Comets and the origin and evolution of life*. Springer, Berlin Heidelberg.
- Wilson, R. (2009). Organic material in micrometeorites: Processes affecting its delivery to planetary environments. Dissertation, The Open University.